

## REFERENCES

- [Abiteboul 97] S. Abiteboul, S. Cluet, V. Christophides, T. Milo, G. Moerkotte, J. Simeon, Querying Documents in Object Databases, *International Journal on Digital Libraries* 1, 1, 1997, pp. 5--19.
- [Adelberg 98] B. Adelberg, NoDoSE - A Tool for Semi-Automatically Extracting Structured and Semistructured Data from Text Documents, *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, 1998, Seattle, Washington, pp.283--294
- [Alexandria Crosswalk] Alexandria Digital Library. Crosswalk: FGDC Content Standards for Digital Geospatial Metadata to USMARC. URL: <http://alexandria.sdc.ucsb.edu/public-documents/metadata/fgdc2marc.html>
- [Ashish 97] N. Ashish and C. Knoblock, Semi-Automatic Wrapper Generation for Internet Information Sources, *Proceedings of the CoopIS'97*, 1997.
- [Ashish 97] N. Ashish, C. Knoblock, Semi-Automatic Wrapper Generation for Internet Information Sources, *Proceedings of the CoopIS'97*, 1997.
- [Atzeni 97a] P. Atzeni, G. Mecca, P. Merialdo, To Weave the Web, *Proceedings of the Twenty-third International Conference on Very Large Data Bases 1997*, Athens, Greece, pp. 206--215.
- [Atzeni 97b] P. Atzeni and G. Mecca, Cut and Paste (*sic*), *Proceedings of the 16th ACM PODS*, 1997, pp. 144--153.
- [Beckett] Dave Beckett, Resource Description Framework (RDF) Resources, <http://www.cs.ukc.ac.uk/people/staff/djb1/research/metadata/rdf.shtml>
- [Blum & Mitchell 98] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, 92-100, 1998.
- [Brickley & Guha 99] D. Brickley, R.V. Guha, Resource Description Framework (RDF) Schema Specification W3C Proposed Recommendation, March 1999 <http://www.w3.org/TR/PR-rdf-schema/>
- [Brin 98] S. Brin, Extracting Patterns and Relations from the World Wide Web, *Proceedings of the WebDB Workshop (at EDBT'98)*, 1998.
- [Chakrabarti et al. 99] S. Chakrabarti, B. E. Dom, S. Ravi Kumar, Prabhakar Raghavan, S. Rajagopalan, A. Tomkins, D. Gibson, and J. Kleinberg, Mining the Web's Link Structure, *IEEE Computer*, Vol 32(8), pp. 60-67, August 1999.
- [Chalmers 93] M. Chalmers, Using a Landscape Metaphor to Represent a Corpus of Documents. In: Frank, A. U., and Campari, I., (eds.), *Spatial Information Theory. A Theoretical Basis for GIS*, vol. 716, *Lecture Notes in Computer Science*, Springer, Berlin, D: 377-390, 1993.
- [Chawathe 94] S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman and J Widom, The TSIMMIS Project: Integration of Heterogeneous Information Sources, *IPSJ Conference 1994*, Tokyo,.
- [Clever 99] Members of the Clever Project, IBM Almaden Center, Hypersearching the Web, *Scientific American*, June 1999.

- [Cowie 96] J. Cowie and W. Lehnert, Information Extraction, Communications of the ACM 39, 1, 1996, pp. 80--91.
- [Craven et al. 98] Mark Craven, Dan DiPasquo, Dayne Freitag, Andrew McCallum, Tom Mitchell, Kamal Nigam, and Sean Slattery. Learning to extract symbolic knowledge from the world wide web. In Proceedings of the 15th National Conference on Artificial Intelligence, 509-516, 1998.
- [Craven 98] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, S. Slattery, Learning to Extract Symbolic Knowledge from the World Wide Web, Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98), 1998, Madison, Wisconsin, pp. 509--516.
- [CSDGM1] International Federation of Library Associations and Institutions Digital Libraries: Metadata Resources [http://www.ifla.org/II/me\[CSDGM1\]](http://www.ifla.org/II/me[CSDGM1])
- [Delcambre 97] L.M.L. Delcambre, D. Maier, R. Reddy, L. Anderson, Structured Maps: Modeling Explicit Semantics Over a Universe of Information, International Journal on Digital Libraries, 1, 1, 1997, pp. 20--35.
- [Delcambre 99] L. Delcambre and D. Maier, Models for Superimposed Information, Proceedings of the Workshop on the World Wide Web and Conceptual Modeling (WWWCM'99), Springer Verlag, LNCS 1727, Editors: P.P. Chen, D.W. Embley, J. Kouloumdjian, S.W. Liddle, J.F. Roddick, Paris, France, 1999, pp. 264--280.
- [Doorenbos 97] R.B. Doorenbos, O. Etzioni, D.S. Weld, A Scalable Comparison-Shopping Agent for the World-Wide Web, Proceedings of the First International Conference on Autonomous Agents, Marina Del Rey, California, 1997, pp. 39--48.
- [Embley 80] D.W. Embley, Programming With Data Frames for Everyday Data Items, Proceedings of the 1980 National Computer Conference, Anaheim, California, pp. 301--305.
- [Fabrikant 99] S.I. Fabrikant, Spatial Metaphors for Browsing large Data Archives, PhD Dissertation, University of Colorado-Boulder, 1999.
- [FGDC] U.S. Federal Geographic Data Committee (FGDC). Content Standards for Digital Geospatial Metadata. URL: <ftp://fgdc.er.usgs.gov/metadata/csdgm/>
- [Freitag 98] Dayne Freitag. Information extraction from HTML: Application of a general machine learning approach. In Proceedings of the 15th National Conference on Artificial Intelligence, 517-523, 1998.
- [Freitag 98] D. Freitag, Information Extraction from HTML: Application of a General Machine Learning Approach, Proceedings of AAAI/IAAI, 1998, pp. 517--523.
- [Goldman et al 95] S. A. Goldman and M. K. Warmuth. Learning binary relations using weighted majority voting. Machine Learning, 20, 245-271, 1995.
- [Goodchild 97] M. F. Goodchild and G. J. Hunter, A Simple Positional Accuracy Measure for Linear Features, Int. Journal of Geographical Information Science, 11, pp. 299-306, 1997.
- [Google] About Google. <http://www.google.com/about.html>

- [Gupta 97] Gupta and V. Harinarayan and A. Rajaraman, Virtual Database Technology, SIGMOD Record 26, Number 4, December 1997, pp. 57--61.
- [Hammer 97] J. Hammer, H. Garcia-Molina, J. Cho, R. Aranha, A. Crespo, Extracting Semistructured Information from the Web, Proceedings of the Workshop on Management of Semistructured Data, 1997, Tucson, Arizona.
- [Harold 99] E. R. Harold, XML Bible, IDG Books Worldwide, 1999  
<http://metalab.unc.edu/xml/books/bible/>.
- [Hill 96] L. Hill, Stocking the Digital Library with Geospatial Information, April 1996,  
[http://alexandria.sdc.ucsb.edu/~lhill/diglib\\_t.html](http://alexandria.sdc.ucsb.edu/~lhill/diglib_t.html).
- [Introna&Nissenbaum 00] Lucas Introna and Helen Nissenbaum, Defining the Web: The Politics of Search Engines, IEEE Computer, 33, 1, January 2000, pp 54-62.
- [Kleinberg 99] J. M. Kleinberg, Authoritative Sources in a Hyperlinked Environment, Jour. of the ACM, Vol 46 (5), pp. 604-632, 1999.
- [Kushmerick 97] N. Kushmerick, D.S. Weld, R. Doorenbos, Wrapper Induction for Information Extraction, Proceedings of the 1997 International Joint Conference on Artificial Intelligence, pp. 729--735,
- [Lang 95] K. Lang. Newsweeder: Learning to filter netnews. In Proceedings of the 12th International Conference on Machine Learning, pages 331-339. 1995.
- [Lehnert 94] W. Lehnert, C. Cardie, D. Fisher, J. McCarthy, E. Riloff and S. Soderland, Evaluating an Information Extraction System, Journal of Integrated Computer-Aided Engineering 1, 6, 1994.
- [Maier 99] D. Maier and L. Delcambre, Superimposed Information for the Internet, Proceedings of the ACM SIGMOD Workshop on the Web and Databases (WebDB'99), Editors S. Cluet and T. Milo, Philadelphia, Pennsylvania, 1999.
- [McCallum et al. 99] Andrew McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. A machine learning approach to building domain-specific search engines. In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, 1999.
- [Muslea 98] I. Muslea, S. Minton, C. Knoblock, STALKER: Learning Extraction Rules for Semistructured, Web-based Information Sources, Proceedings of AAAI'98: Workshop on AI and Information Integration, Madison, Wisconsin, 1998.
- [Nigam et al 98] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell. Learning to classify text from labeled and unlabeled documents: A general machine learning approach. In Proceedings of the 15th National Conference on Artificial Intelligence, 792-799, 1998.
- [Nigam et al. to appear] Kamal Nigam, Andrew McCallum, Sebastian Thrun and Tom Mitchell. Text classification from labeled and unlabeled documents using EM. Machine Learning, to appear.
- [Quinlan 90] J. R. Quinlan. Learning logical definitions from relations. Machine Learning, 5, 239-266, 1990.

- [Rabiner 89] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286, 1989.
- [Revesz 98] Kanjamala, P., Revesz, P.Z., Wang, Y., MLPQ/GIS: A Geographic Information System using Linear Constraint Databases, *Proc. Ninth International Conference on Management of Data*, pp. 389-392, Hyderabad, India, December 1998.
- [Sahuguet 99] A. Sahuguet and F. Azavant, Looking at the Web through XML glasses, *Proceedings of the Fourth International Conference on Cooperative Systems (CoopIS'99)*, 1999, Edinburgh, Scotland, UK,
- [Schapire & Singer a] Robert E. Schapire and Yoram Singer. BoosTexter: A boosting-based system for text categorization. *Machine Learning*, to appear.
- [Schapire & Singer b] Robert E. Schapire and Yoram Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, to appear. Early version in COLT '98.
- [Smith 96] T. R. Smith, The Meta-Information Environment of Digital Libraries, *D-Lib Magazine*, July/August 1996, <http://www.dlib.org/dlib/july96/new/07smith.html>.
- [Smith 97] D. Smith and M. Lopez, Information Extraction for Semi-structured Documents, *Proceedings of the Workshop on Management of Semistructured Data*, 1997, Tucson, Arizona.
- [Soderland 97], S. Soderland, Learning to Extract Text-based Information from the World Wide Web, *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, 1997, Newport Beach, California, pp. 251--254.
- [Srihari 92] S.N. Srihari, High-performance reading machines, *Proceedings of IEEE*, 80(7), 1120-1132, July 1992.
- [WordNet 98] WordNet: An Electronic Lexical Database, Christiane Fellbaum (Ed.), MIT Press, 1998.
- [XML] Extensible Markup Language (XML), <http://www.w3.org/XML/>.